Assignment 1, Liner Regression

2111512017 李疆

- 数据读取(Load Data)
- 数据归一化(features Normalize)
- 训练梯度下降 (Training - Gradient Descent)
- 结果评估(Estimate the model)

数据读取(Load Data)

OSX系统下，matlab实现

选取肺活量，短跑，跳远作为特征，拟合估计长跑成绩。

```
[digit, txt, raw] = xlsread('psydata.xlsx'); %?表格已处理过，把跑步成绩文本信息，转化为数字，添加新列保存，便于读取数据。男女长跑成绩都化为800米跑步时间。

%选取肺活量，短跑，跳远作为特征arr_X，拟合长跑成绩
trainSize = 2000; % Training sample size
testSize = size(digit,1) - trainSize;
X = digit(1:trainSize, 4:6);
Y = digit(1:trainSize, 12);
```

数据归一化(eatures Normalize)

`function [ X_normal, mu, sigma ] = featureNormalize( X )`

求出样本均值mu和方差sigma，处理输入`x(i) = (x(i) - mu ) / sigma`.

添加一列x(0) = 0;

训练梯度下降 (Training - Gradient Descent)

设置学习率

$$\alpha = 0.01$$

, 迭代次数 `num_iters = 2000`

- 代价损失函数 cost function

```
J = sum((X * theta - Y) .^ 2) / ( 2 * m );
```

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

- 梯度下降，每次迭代更新

$$\theta$$

向量，并记录代价函数值

$$J(\theta)$$

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \quad \text{(simultaneously update } \theta_j \text{ for all } j).$$

```
theta = theta - (alpha / m ) * X' * (X * theta - Y);
J_history(iter) = computeCostMulti(X, Y, theta);  %save the cost J
in every iteration
```

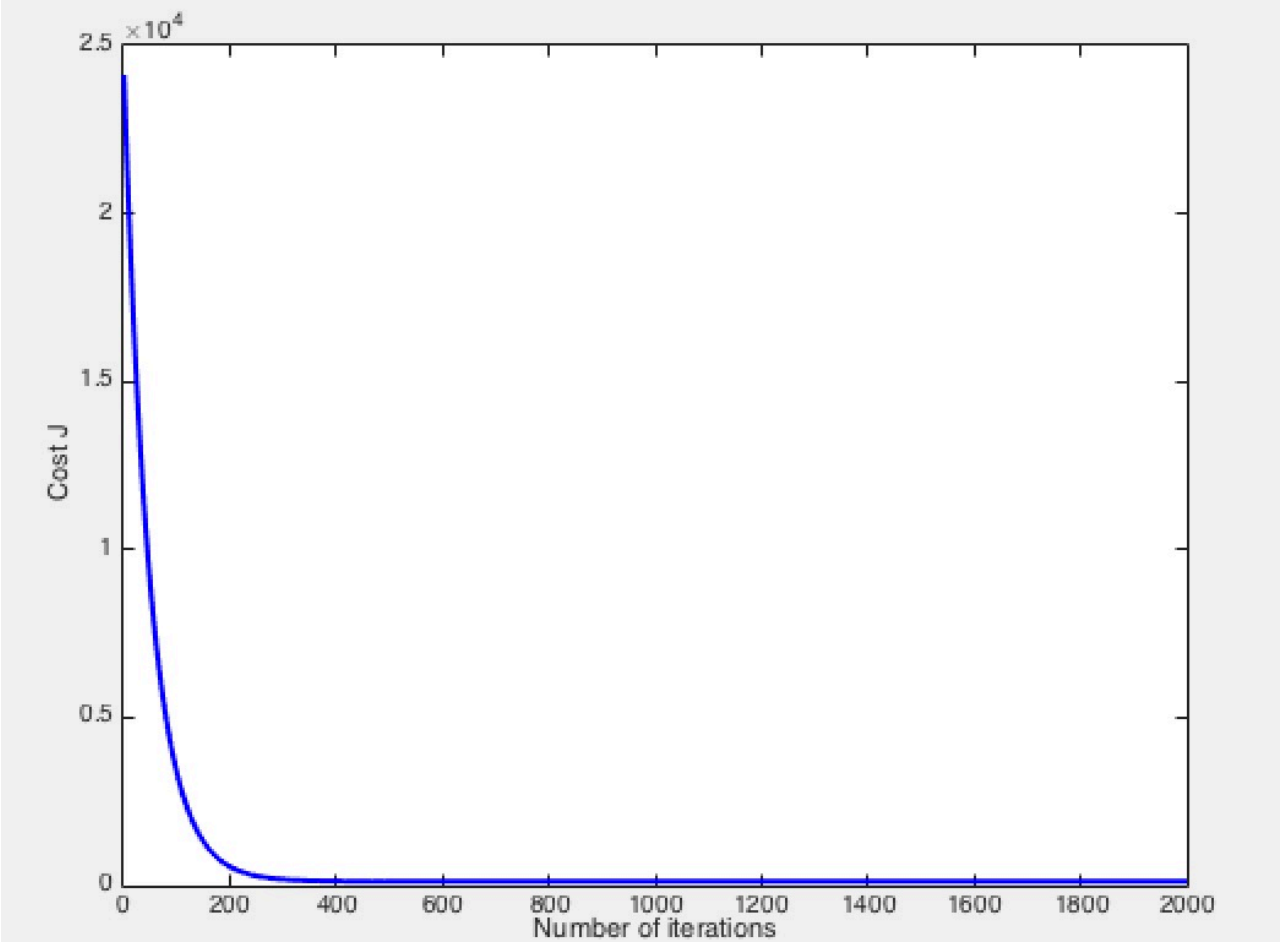
结果评估(Estimate the model)

- 结果输出如下

```
Loading data ...
First 10 examples from the dataset:
x = [3271 8 186], y = 227
x = [3684 8 197], y = 229
x = [2229 8 200], y = 210
x = [2349 9 190], y = 222
x = [2708 9 160], y = 241
x = [2714 9 178], y = 229
x = [2766 9 185], y = 237
x = [2920 10 150], y = 253
x = [4531 7 230], y = 202
x = [3270 9 148], y = 259
Normalizing Features ...
Running gradient descent ...
Theta computed from gradient descent:
 220.323400
 -0.249902
 15.106152
 -6.515838

accurate rate: 206
rate = sum((predict_y-y).^2) / length(y)
```

- 由下图可知，模型收敛



- 对测试数据评估误差

  令

$$testError = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2$$

```
acc_rate = sum( (X_test * theta - Y_test) .^ 2 ) / m;
```

得到

$$testError = 206$$

（男女长跑成绩统一表示为 秒 / 800米 ）